

A large-scale *in vivo* analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly

Shijia Chen, Grigorios Oikonomou, Cindy N. Chiu, Brett J. Niles, Justin Liu, Daniel A. Lee, Igor Antoshechkin and David A. Prober*

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Received October 23, 2012; Revised and Accepted December 5, 2012

ABSTRACT

Zinc-finger nucleases (ZFNs) and TAL effector nucleases (TALENs) have been shown to induce targeted mutations, but they have not been extensively tested in any animal model. Here, we describe a large-scale comparison of ZFN and TALEN mutagenicity in zebrafish. Using deep sequencing, we found that TALENs are significantly more likely to be mutagenic and induce an average of 10-fold more mutations than ZFNs. We observed a strong correlation between somatic and germ-line mutagenicity, and identified germ line mutations using ZFNs whose somatic mutations rates are well below the commonly used threshold of 1%. Guidelines that have previously been proposed to predict optimal ZFN and TALEN target sites did not predict mutagenicity *in vivo*. However, we observed a significant negative correlation between TALEN mutagenicity and the number of CpG repeats in TALEN target sites, suggesting that target site methylation may explain the poor mutagenicity of some TALENs *in vivo*. The higher mutation rates and ability to target essentially any sequence make TALENs the superior technology for targeted mutagenesis in zebrafish, and likely other animal models.

INTRODUCTION

Zinc-finger nucleases (ZFNs) and TAL effector nucleases (TALENs) have recently emerged as powerful tools for generating targeted genomic mutations. These proteins bind specific DNA sequences and induce double-strand DNA breaks that are repaired by non-homologous end joining (1), an error-prone process that often results in insertion or deletion (indel) mutations. ZFNs have been

studied for several years, but their widespread use has been limited by the difficulty of targeting them to specific DNA sequences. Selection assays for identifying zinc fingers that bind specific targets are laborious and challenging for non-specialist laboratories (2,3). An alternative method, known as modular assembly, combines pre-selected zinc-finger modules into arrays (1). These ZFNs are relatively easy to generate but have low success rates (4), although significant progress has recently been made (5,6). Proprietary methods have also been used to generate ZFNs that are effective in zebrafish (7), but these nucleases must be purchased and are expensive. Another approach, context-dependent assembly (CoDA) (8), does not require selection assays and was claimed to have a success rate comparable with selection-based methods. However, the sequences that can be targeted using CoDA are limited (8) and, as for all ZFN technologies, there is no established code for specific zinc finger/DNA interactions. In contrast, TALENs contain a variable number of repeated modules that each preferentially binds a specific nucleotide. Therefore, TALENs can in principle be targeted to any DNA sequence without the need for selection assays. Furthermore, TALENs can be constructed using standard molecular biology techniques (9). Both ZFNs and TALENs can induce mutations in zebrafish (3,5–15), but these approaches have not been extensively tested and compared for mutagenicity in any animal model. Here, we describe a large-scale analysis and comparison of ZFN and TALEN mutagenicity in developing zebrafish embryos.

MATERIALS AND METHODS

Animal care

All experiments were performed by mating TL and AB wild-type zebrafish strains using standard protocols (16) in accordance with the California Institute of Technology Institutional Animal Care and Use Committee guidelines.

*To whom correspondence should be addressed. Tel: +1 626 395 3071; Fax: +1 626 395 2211; Email: dprober@caltech.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2013. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Construction of ZFNs and TALENs

ZFNs and TALENs were designed using the ZiFIT Targeter (<http://zifit.partners.org/ZiFiT>) (9,17). DNA fragments encoding zinc-finger arrays were synthesized (Epoch Life Science Inc.) and cloned into FokI EL/KK (18) heterodimeric expression vectors by BamHI/XbaI (pMLM290 and pMLM292) or NotI/XbaI (pMLM800 and pMLM802) digestion (19). Each ZFN contained three zinc fingers. TALE repeat arrays were constructed using the REAL Assembly TALEN Kit (9) and were cloned into the wild-type FokI expression vectors JDS70, JDS71, JDS74 and JDS78. All TALENs were sequence verified before mRNA synthesis. Plasmids were obtained from the non-profit plasmid repository Addgene.

RNA transcription and injection

ZFN and TALEN expression plasmids were linearized with PmeI and purified using the polymerase chain reaction (PCR) Purification Kit (Qiagen). mRNA was synthesized using 500 ng of purified linear DNA as template and the mMessage mMachine T7 Ultra kit (Ambion). The transcription reaction yielded ~20 µg of polyA tailed mRNA, which was dissolved in 20 µl of nuclease-free water. mRNA synthesis and polyA tailing were verified by agarose gel electrophoresis. Final mRNA concentrations ranged from 0.8 to 1.2 µg/µl. Approximately 50–100 pg of each ZFN or TALEN mRNA was injected into the cell of zebrafish embryos at the one-cell stage. mRNA concentrations that were sufficient to cause developmental defects in 10–50% of injected embryos were used to assay for somatic mutations and to generate germ line mutants.

Analysis of somatic mutations

For each ZFN and TALEN, genomic DNA was prepared from 12 injected embryos at 72 h post-fertilization (h.p.f.) as previously described (19). Embryos were incubated in 500 µl of sodium dodecyl sulfate lysis buffer [10 mM of Tris pH 8.0, 200 mM of NaCl, 10 mM of ethylenediaminetetraacetic acid (EDTA), 0.5% of sodium dodecyl sulfate, 100 µg/ml of proteinase K] overnight at 50°C with occasional gentle mixing until no clumps were visible. Genomic DNA was then purified by phenol/chloroform extraction and dissolved in 40 µl of TE (10 mM of Tris pH 8.0, 0.5 mM of EDTA). Targeted genomic regions were amplified using Ampliqaq (Invitrogen), with amplicons ranging in size from 175 to 350 bp. Because of the short Illumina sequence read length, for each PCR reaction, one primer was designed to anneal 6–12 bp away from the spacer. PCR products were purified using the PCR Purification Kit (Qiagen) and pooled at roughly equal molar ratios. Sequencing libraries were prepared by following the Illumina TruSeq Genomic DNA protocol without the DNA fragmentation step. The pooled PCR products were end repaired, A-tailed and ligated to TruSeq single-index adaptors. The adaptor-ligated DNA was gel purified and PCR amplified to produce finished libraries. The libraries were sequenced using an Illumina GAIIx machine in the single read

38-nt mode, producing 34.1 million reads, and using an Illumina HiSeq2000 machine in the single-read 50-nt mode, producing 144.1 million reads.

Detecting indels using short-read data is challenging because the commonly used alignment algorithms developed for high-throughput sequencing data, such as bowtie (20) and ELAND, cannot map discontinuous reads. Aligners that use the split-read method, which maps defined portions of the read separately before creating the final alignment, such as tophat (21) and ELAND2, suffer from poor sensitivity of indel detection because of the requirement for the indel site to fall between the mapped portions of the read. Custom implementations of the split-read method suffer from similar sensitivity problems. Also, many split-read aligners, including tophat, were designed for the alignment of RNA-Seq data and are much more sensitive in aligning discontinuous reads that span splice junctions, which is not the case for our data.

To detect indels rigorously, we used a combination of two methods. We first aligned reads to target genomic regions using the SHRiMP2 software package (22,23), which allowed us to identify small indels with high sensitivity. SHRiMP2 uses the vectorized Smith–Waterman algorithm for local alignment during the candidate mapping location identification phase, followed by the full Smith–Waterman alignment to detect single-nucleotide polymorphisms and indels. It has been shown to have the highest sensitivity of the currently available short-read aligners (22,23). Despite its high sensitivity for small indels, SHRiMP2 is unable to map reads across large insertions or deletions. To identify such events in our data, we mapped reads that failed to align with SHRiMP2 using BLAT, which is capable of finding regions of high similarity separated by large gaps (24). Mappings produced by SHRiMP2 were output in the SAM format, whereas the BLAT-native psl format was converted to SAM using the psl2sam.pl script provided by the samtools package (25). The SAM format-defined CIGAR representation of the alignment for each read was then used to identify insertions and deletions using a custom perl script (available on request). We then removed indels that were closer than 17 nt, the length of the shortest PCR primer, to the 5'-end of the read, or did not have at least a 5-nt continuous match on the 3'-end, which may correspond to incorrect alignments. We also filtered out 1-nt indels, which may result from PCR or sequencing errors. The filtered indels produced by SHRiMP and BLAT were merged to produce the final indel list.

Isolation of germ line mutants

Zebrafish embryos injected with a ZFN or TALEN pair were raised to adulthood and mated to other potential founders or wild-type fish. Depending on the somatic mutation rate, genomic DNA was isolated from a pool of 1–6 embryos at 72 h.p.f., with up to 96 embryos tested for each fish. Embryos were incubated in embryo lysis buffer (10 mM of Tris–HCl pH 7.5, 1 mM of EDTA, 50 mM of KCl, 0.3% of Tween 20, 0.3% of NP40) for 10 min at 98°C. Proteinase K was then added to a final

concentration of 1 mg/ml, and the lysis reaction was incubated overnight at 55°C, followed by 10 min at 98°C. One microliter of this solution was used as template for PCR.

Targeted genomic regions were amplified using Amplitaq (Invitrogen) with amplicons ranging in size from 200 to 550 bp. In cases where an indel was expected to delete a restriction enzyme site, half of the PCR reaction was digested with the appropriate restriction enzyme and analysed by agarose gel electrophoresis. In cases where the targeted region did not overlap with a restriction enzyme site, PCR was performed using a standard forward primer and a reverse primer fluorescently labelled with 6-FAM, HEX (Integrated DNA Technologies) or NED (Applied Biosystems). For PCR using fluorescently labelled primers, we performed a final 1-h incubation at 60°C to ensure that an extra adenosine was added to the 3'-end of all PCR products. Fluorescent PCR products were run on an ABI 3730 DNA analyzer (Applied Biosystems), and PCR product sizes were analysed using Peak Scanner (Applied Biosystems). To confirm indel sequences, genomic DNA from single embryos was amplified and sub-cloned using the Strataclone PCR Cloning Kit (Stratagene), and DNA from several independent colonies was sequenced. DNA sequences were analysed using SeqBuilder and SeqMan (DNASar Lasergene).

Statistical methods

Mutation rate and size data do not follow a normal distribution; therefore, we used non-parametric methods to analyse these data. The non-parametric Wilcoxon rank-sum test (also known as Mann–Whitney U-test or Mann–Whitney–Wilcoxon test) was used to examine statistically significant differences in measurement variables (mutation rate or size) between two nominal variables (ZFN versus TALEN). Correlations were used to test whether pairs of variables co-vary. Pearson's linear correlation was used to test interval data, whereas Spearman's rank correlation was used to test ordinal data. Results of the correlations are reported as the r^2 -value (coefficient of determination) and P -value. Data are displayed as scatter or box plots, and significant linear correlations were additionally displayed with a regression line. All tests were two-tailed, and alpha level was $P < 0.05$.

RESULTS

Few CoDA ZFNs induce somatic mutations *in vivo*

In a previous study, 12 of 24 (50%) CoDA ZFNs induced somatic mutations in zebrafish, with indel rates between 1 and 17% (8). However, these ZFNs were pre-screened for activity using a bacterial reporter assay. Because only ~75% of CoDA zinc fingers were found to be active in the reporter assay (8), the success rate of CoDA ZFN pairs that are not pre-screened was estimated at ~28% (50% × 75% left ZFN × 75% right ZFN). Another study found that 3 of 17 (18%) CoDA ZFNs that were not pre-screened induced somatic mutations in zebrafish, with indel rates of 1–3% (11). To more comprehensively

evaluate CoDA ZFN mutagenicity *in vivo*, we generated 84 ZFN pairs targeting 66 zebrafish genes. We screened for indels using deep sequencing, generating an average of 1 200 000 reads per target. In contrast, most studies have analysed ≤96 sequence reads; hence, ZFNs that induced indels at rates <1% would likely not have been identified as mutagenic.

Of the 84 ZFN pairs tested, 21 (25%) induced indels in >1% of sequence reads, in close agreement with the predicted rate of 28% (8,11), and only 5 (6%) produced indel rates >10% (Figure 1c, Supplementary Figure S1 and Supplementary Table S1). Thus, CoDA ZFNs have relatively low success rates and few CoDA ZFNs induce mutations at high rates. A genomic region that was not targeted by a ZFN gave an indel rate of 0.009% (Supplementary Table S1), indicating a false-positive rate of ~0.01%. Surprisingly, many ZFNs induced indels at low frequencies (Figure 1c, Supplementary Figure S1 and Supplementary Table S1). Fifty-four ZFNs (64%) induced indels at rates between the negative control value of 0.01% and 1%, and 18 (21%) induced indels at rates between 0.1% and 1%. Only nine (11%) produced indels at frequencies below the negative control. Thus, most CoDA ZFNs are mutagenic *in vivo* but induce mutations at frequencies below the commonly used threshold of 1%. A caveat to using Illumina-based deep sequencing for indel detection is that one primer must be located close to the targeted region because of the short-sequence read length. Therefore, deletions removing more than ~9 bp beyond the spacer on one side of the targeted region will not be detected. Indeed, of the 20 ZFN-induced germ line mutations in 8 genes that we confirmed using Sanger sequencing (see below), 2 would not have been detected using deep sequencing, suggesting that our methodology may underestimate indel rates by ~10%, although the sample size is modest. An additional caveat is that our requirement that both the 5'- and 3'-ends of a sequence read align to its reference sequence (see 'Materials and Methods' section) will result in a failure to detect large insertions. However, none of the 91 germ line mutations that we identified (see below) contain insertions that are too large to align to their reference sequence using our approach, suggesting that this is unlikely to significantly affect quantification of somatic mutation rates.

Germ line mutations can be generated using ZFNs that have low somatic indel rates

To determine the relationship between somatic and germ line mutation rates, we isolated germ line mutants using ZFNs that induced somatic mutations at a range of frequencies. We observed a strong correlation between somatic and germ line mutation rates (Figure 2, $r^2 = 0.92$, $P = 1.8 \times 10^{-4}$ for eight ZFN pairs). Surprisingly, ZFNs with somatic indel rates as low as 0.27% and 0.33% produced germ line mutations in 8% (3/38) and 7% (3/42) of fish, respectively. Thus, ZFNs whose somatic indel rates are well below the commonly used threshold of 1% can be used to isolate germ line mutants at reasonable frequencies. As a result, it should be possible to isolate germ line mutations using at least

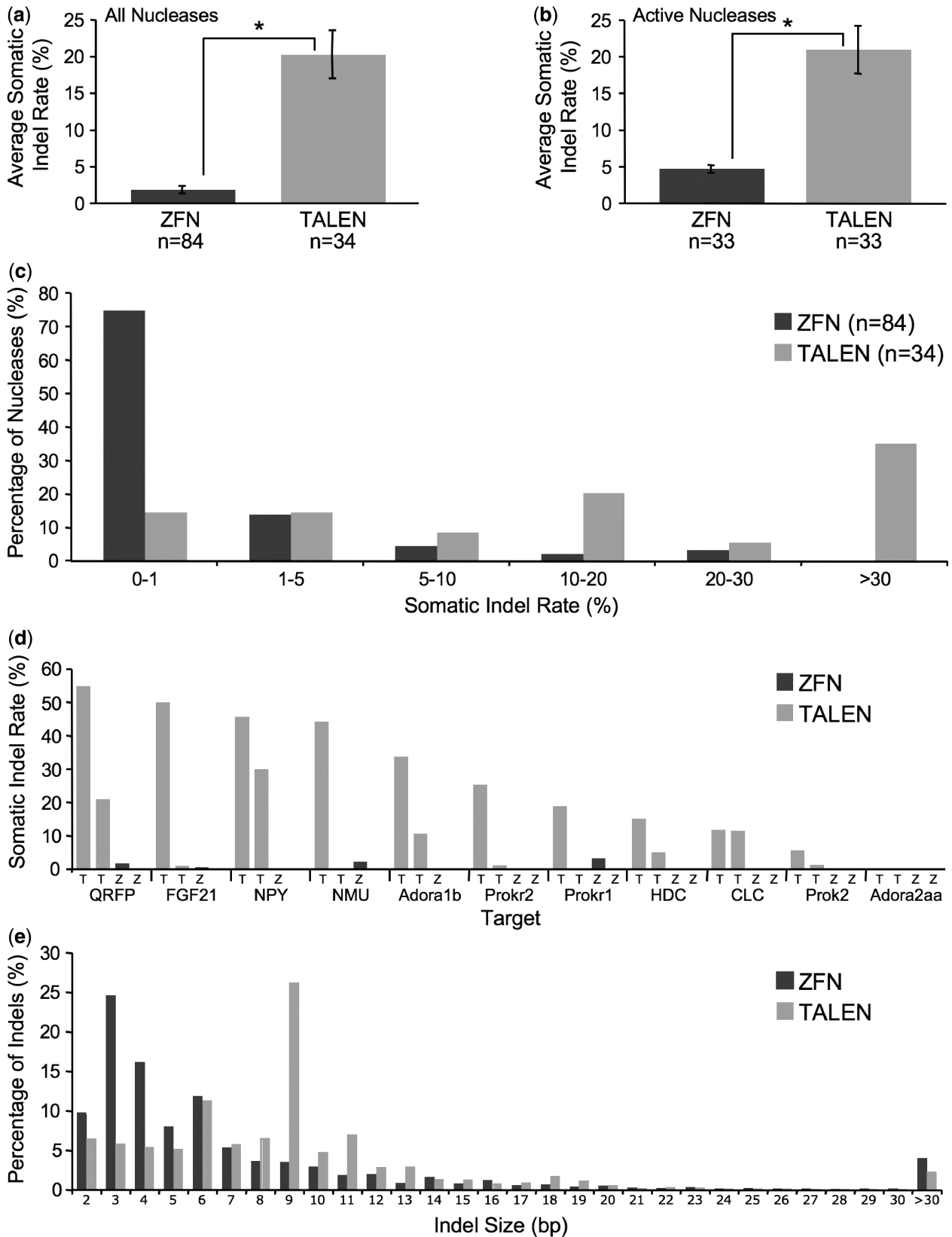


Figure 1. TALENs induce somatic indels at higher rates than ZFNs in zebrafish embryos. Average somatic indel rates are shown for all 84 ZFN and 34 TALEN pairs tested (a) and for the 33 ZFN and 33 TALEN active pairs (b). Nucleases that induce somatic indels at rates >0.27% are defined as active, because this rate is sufficient to generate germ line mutations. The difference in mutation rates for ZFNs and TALENs is statistically significant, indicated by asterisks, with $P = 5.1 \times 10^{-12}$ for all nucleases (a) and $P = 6.0 \times 10^{-5}$ for active nucleases (b) using the Wilcoxon

(continued)

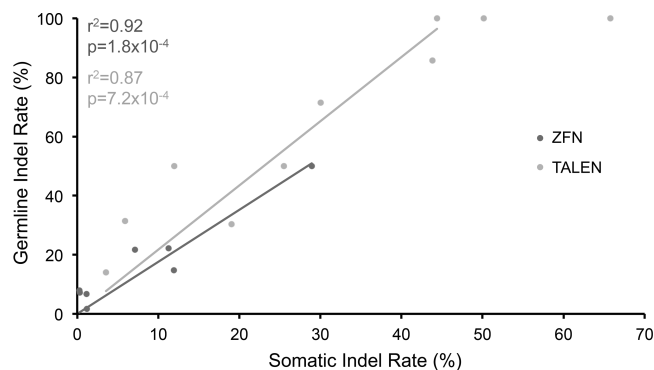


Figure 2. Relationship between rates of somatic and germ line mutations. There is a significant correlation between rates of somatic and germ line indels for ZFNs ($r^2 = 0.92$, $P = 1.8 \times 10^{-4}$, $n = 8$ ZFNs using Pearson's correlation) and TALENs ($r^2 = 0.87$, $P = 7.2 \times 10^{-4}$, $n = 8$ TALENs using Pearson's correlation). The linear regression lines are shown. Three TALENs induced germ line mutations in 100% of injected fish. Among these three TALENs, only the TALEN with the lowest somatic mutation rate was used in the Pearson's correlation calculation and to generate the linear regression line to avoid a ceiling effect.

39% (33/84) of the ZFNs that we tested (Supplementary Table S1); we refer to these ZFNs as active. Because our false-positive rate is 0.01%, ZFNs with somatic indel rates $<0.27\%$ might be useful for generating germ line mutations. For all mutations analysed, the number of mutant embryos produced by each founder was $<50\%$, indicating that mutations were present in a subset of founder germ cells (Supplementary Table S2a). Germ line indels ranged in size from a 5-bp insertion to a 61-bp deletion (Figure 3). ZFNs induced similar somatic and germ line indel sizes regardless of mutation rate (Figure 3 and Supplementary Figure S2a).

Most TALENs induce somatic mutations *in vivo*

TALEN technology has recently been applied to zebrafish (9,11–15). We sought to comprehensively assess the use of TALENs for zebrafish mutagenesis and compare them to ZFNs. We used the REAL method (9) to generate 34 TALEN pairs that target 18 genes. Twenty-one (62%) of the TALENs induced somatic indels at rates $>10\%$, 29 (85%) at rates $>1\%$ and all at rates $>0.1\%$ (Figure 1c, Supplementary Figure S1 and Supplementary Table S3). 33/34 (97%) induced somatic indels at rates $>0.27\%$, and should therefore generate germ line mutations at reasonable frequencies. The single TALEN pair below this threshold had an indel rate of 0.16%, which is still above the false-positive rate of 0.01%. We conclude that most, if not all, TALENs can induce somatic mutations, and in this respect TALENs are superior to CoDA ZFNs.

A potential caveat to this conclusion is that we compared ZFNs containing FokI EL/KK heterodimers with TALENs containing FokI homodimers. However, this difference is unlikely to underlie the higher mutagenicity observed for TALENs, as TALENs containing FokI heterodimers exhibit similar or higher mutation rates than those containing homodimers (13), and the EL/KK heterodimers used in our study and the ELD/KKR FokI heterodimers primarily used by Cade *et al.* (13) induce mutations at similar rates in zebrafish (11). Nevertheless, ELD/KKR heterodimers have been shown to be more active than EL/KK heterodimers in some contexts (26), and they might partially account for the difference in ZFN and TALEN mutation rates in our study.

Somatic and germ line mutation rates are similarly correlated for ZFNs and TALENs

We isolated germ line mutations using several TALEN pairs and observed a strong correlation between somatic and germ line mutation rates (Figure 2, $r^2 = 0.87$, $P = 7.2 \times 10^{-4}$ for eight TALEN pairs). Notably, the slopes of the linear regression lines and correlation coefficients are similar for TALENs and ZFNs, indicating that the relationship between somatic and germ line mutation rates is similar for ZFNs and TALENs. TALEN-induced germ line indels ranged in size from a 23-bp insertion to a 203-bp deletion (Supplementary Figure S3 and Supplementary Table S2b). TALENs induced similar somatic and germ line indel sizes regardless of mutation rate (Supplementary Figures S2b and S3).

Mutagenic TALENs exhibit higher indel rates than mutagenic CoDA ZFNs

Having found that TALENs are much more likely than CoDA ZFNs to be active (i.e. somatic indel rate $>0.27\%$), we next compared their somatic mutagenicity. The average somatic indel rates for the 84 ZFNs and 34 TALENs were 2% and 20%, respectively (Figure 1a and Supplementary Figure S1). If we only compare active nucleases, the indel rates were 5% and 21%, respectively (Figure 1b). These differences are statistically significant ($P = 5.1 \times 10^{-12}$ for all nucleases, $P = 6.0 \times 10^{-5}$ for active nucleases) and result from a shift in the distribution of indel rates for TALENs compared with ZFNs and higher rates for the most mutagenic TALENs (Figure 1c and Supplementary Figure S1). For 10/11 genes that were targeted using both ZFNs and TALENs, at least one TALEN pair was more active than a ZFN pair, in most cases by a large margin (Figure 1d and Supplementary Table S4). Therefore, not only are TALENs more likely than CoDA ZFNs to be active, they also induce mutations

Figure 1. Continued

rank-sum test. (c) Distribution of somatic indel rates for ZFNs and TALENs. Most ZFNs induced somatic indels at frequencies $<1\%$, whereas most TALENs induced indels at significantly higher rates. (d) Somatic indel rates for 11 genes that were targeted with one or two pairs of ZFNs and TALENs. For 10/11 genes, a TALEN pair induced higher indel rates than a ZFN pair. Z and T indicate ZFN and TALEN data. See Supplementary Table S4 for indel rate values. (e) Distribution of somatic indel sizes for ZFNs and TALENs. TALEN-induced indels were significantly larger than ZFN-induced indels, with $P = 2.2 \times 10^{-16}$ using the Wilcoxon rank-sum test. Median indel sizes were 4nt and 9nt for ZFNs and TALENs, respectively. Data are based on 5527940 and 2893241 sequence reads containing indels induced by TALENs and ZFNs, respectively.

```

ENSDARG00000035732 Arnt11b
Germline indel rate=50.0%
Somatic indel rate=28.9%
WT:  GGCATGGACTTCAACCGCAAAGAGGAAGGGCAGCAT
MT1:  GGCATGGACTTCAAGAGACATGCAAGAGGAAGGGC +5 (-2 and +7)
MT2:  GGCATGGACTTCAACCAGCAAAGAGGAAGGGCGAA +2
MT3:  GGCATGGACTTCAAC---AAGAGGAAGGGCAGCAT -3
MT4:  GGCATGGACTTCAA---GAGGAAGGGCAGCAT -6

ENSDARG00000006791 Arnt11a
Germline indel rate=22.2%
Somatic indel rate=11.3%
WT:  GGCATGGACTACAACCGCAAAGAGGAAGGGCAGCAC
MT1:  GGCATGGACTACA---AGGAAGGGCAGCAC -8
MT2:  GGCATGGACTACAACC-----> -27
MT3:  GGCATGGACTACAA-----> -61

ENSDARG00000075694 Adorala
Germline indel rate=21.7%
Somatic indel rate=7.1%
WT:  ATGGTCTACTTCAACTTCTTCGGCTGGGTGCTTCC
MT1:  ATGGTCTACTTCAACTTCAAGATTCGGCTGGGTGC +4
MT2:  ATGGTCTACTTCAAC---TTCGGCTGGGTGCTTCC -3
MT3:  ATGGTCTACTTCAACT-----GCTGGGTGCTTCC -6
MT4:  ATGGTCTACTTCAACTTC---GGCTGGGTGCTTCC -3

ENSDARG000000068422 Gpr103b
Germline indel rate=14.7%
Somatic indel rate=11.9%
WT:  CTTTACACCACTTCATAATGGTGGCGCTGTTCCCT
MT1:  CTTTACACCAACC---FAATGGTGGCGCTGTTCCCT -4
MT2:  CTTTACACCAAC---C---TGGTGGCGCTGTTCCCT -6 (-7 and +1)
MT3:  CTTTACACCACTTCG---TGGTGGCGCTGTTCCCT -2 (-4 and +2)

ENSDARG000000036222 NPY
Germline indel rate=7.9%
Somatic indel rate=0.3%
WT:  ACAAAAGCCGACAACCCGGGAGAGGACGCCA CCTGC
MT1:  ACA-----AGAGGACGCCA CCTGC -17
MT2:  ACAAA-----AGAGGACGCCA CCTGC -15

ENSDARG000000094637 QRFP
Germline indel rate=7.1%
Somatic indel rate=0.3%
WT:  CAGACCACAGTCTTCTTGTGGTGCTACTGGT
MT1:  CAGACCACAGTCTTCTTCTTGTGGTGCTACTAC +4

ENSDARG000000057239 Tph2
Germline indel rate=6.7%
Somatic indel rate=1.2%
WT:  AGAGAGGACAACATCCACAGCTGGAGGACGTGTC
MT1:  AGAGAGGACAACAT---CAGCTGGAGGACGTGTC -4
MT2:  AGAGAGGACAACATCGGGGACAGCTGGAGGACG +4 (-2 and +6)

ENSDARG000000069446 DBH
Germline indel rate=1.7%
Somatic indel rate=1.2%
WT:  GAGCATCCCATCCTATCGTTGCATGAGCTCAATAT
MT1:  GAGCATCCCATCCTATCGTTGCATGAGCTCA +4

```

Figure 3. Sequences of ZFN-induced germ line mutations. ZFN target sequences and spacer sequence are highlighted in yellow and grey, respectively. Deletions are indicated by red dashes and insertions are highlighted in blue. Only mutations that were analysed using Sanger sequencing are shown.

at higher rates. The distribution of indel sizes is also significantly different for ZFN- and TALEN-induced mutations ($P = 2.2 \times 10^{-16}$), with median indel sizes of 4nt and 9nt, respectively (Figure 1e).

Published guidelines do not predict ZFN or TALEN mutagenicity *in vivo*

Several guidelines have been proposed to select optimal ZFN and TALEN target sites but have not been extensively tested in an animal. We used our data set to evaluate how well these guidelines predict success in generating

mutations *in vivo*. First, we used the confidence score provided by the ZiFIT Targeter for ZFNs that are generated using the OPEN method (27). This score is based on analysis of zinc fingers that activate transcription of a reporter gene in a bacterial assay, which is correlated with ZFN activity (2,10,28–30). Confidence scores range from 0 to 9, with 9 indicating the greatest likelihood of mutagenicity. We compared scores for the 58 CoDA ZFN pairs that we tested that could have been generated using the OPEN method, but found no correlation between confidence score and indel rate (Figure 4a and b, $P = 0.69$ and $P = 0.65$ for average and lowest score of ZFN pairs). Second, it has been suggested that ZFNs are more likely to be mutagenic if most or all nucleotide triplets in the target sequence start with a guanine (3,4,31–36). We found no correlation between mutation rate and target sites containing four, five or six Gxx triplets (Figure 4c, $P = 0.30$). ZFN targets containing only four Gxx triplets might have lower mutation rates, but our data set contains few of these cases, and the difference that we observed is not statistically significant (Figure 4c). However, we did observe a significant negative correlation between the target sequence spacer length and mutagenicity (Figure 4d, $P = 0.005$). ZFN targets with a 7-bp spacer had much lower mutation rates than those with 5- or 6-bp spacers.

We also evaluated guidelines proposed for designing mutagenic TALENs. First, Cermak *et al.* (37) proposed guidelines based on TAL effectors found in nature. In addition to the well-established requirement that TALEN binding sites should be preceded by a T (Guideline 1) (37–39), which we followed for all TALENs tested, they suggest that TALEN binding sites should not have a T at position 1 (guideline 2) or an A at position 2 (guideline 3) and should have a T at the last position (guideline 4). They also suggest that TALEN targets should have a nucleotide composition within 2 standard deviations of the average nucleotide composition of natural TALE targets (guideline 5). We failed to detect correlations between TALEN mutagenicity and any of these guidelines (Figure 4e–i and Supplementary Table S5, $P > 0.16$ for all guidelines). Second, Streubel *et al.* (40) generated artificial TALE arrays and tested their ability to activate a reporter gene in plant cells. Based on their results, they proposed that TALENs should contain at least three to four repeats that bind C or G, whereas stretches of repeats that interact with A or T should be avoided, especially at the ends of target sites. We found no correlation between indel rate and target G+C content (Figure 4j, $r^2 = 0.07$, $P = 0.12$), although none of our targets contained fewer than four G+C nucleotides. We also did not observe a relationship between indel rates and A/T repeats. Two TALEN target sites contained stretches of six or seven A/T nucleotides, yet the TALENs targeting these sites exhibited high-somatic mutation rates (Supplementary Table S5). We note, however, that the guidelines suggested by Streubel *et al.* are based on monomeric TALE proteins and may have less impact in the context of TALEN dimers. Finally, we found that diverse sequences can serve as nuclease targets (Supplementary Figure S4), similar to observations in human cells (41). We conclude that these guidelines have

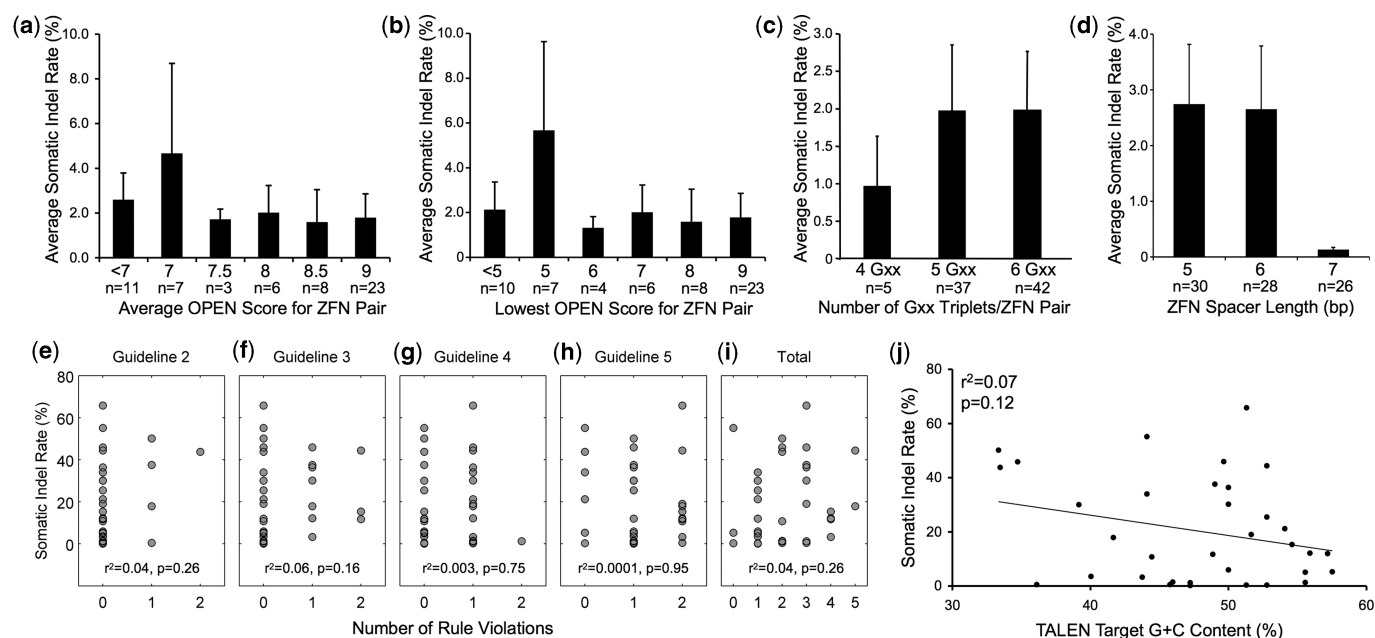


Figure 4. ZFN and TALEN targeting guidelines do not predict mutagenicity in zebrafish embryos. There is no correlation between somatic indel rate and the average (a) or lowest (b) OPEN score for each ZFN pair [$r^2 = 0.003$, $P = 0.69$ for (a) and $r^2 = 0.004$, $P = 0.65$ for (b) using Spearman's rank correlation]. Data for 58 ZFN pairs, for which ZiFiT OPEN scores are available, are shown. (c) There is no correlation between somatic indel rate and the presence of four, five or six Gxx triplets in the ZFN target sequence ($r^2 = 0.01$, $P = 0.30$ using Spearman's rank correlation). ZFN targets containing four Gxx triplets may have lower mutation rates but do not reach a significant threshold. (d) There is a correlation between mutation rate and ZFN spacer length ($r^2 = 0.09$, $P = 0.005$ using Spearman's rank correlation). ZFN targets containing 7-bp spacers have lower mutation rates than those containing 5- or 6-bp spacers. There is no significant difference in mutation rates for ZFN targets containing 5- or 6-bp spacers ($P = 0.42$ using the Wilcoxon rank-sum test). For (a–d), n indicates number of ZFN pairs. Panels (e–i) show somatic indel rates for each TALEN pair for which neither, one or both target half-sites violate one of four design guidelines: (e) no T at position 1, (f) no A at position 2, (g) T at last position and (h) nucleotide composition within 2 standard deviations of natural TALE targets (A = 0 to 63%, C = 11 to 63%, G = 0 to 25%, T = 2 to 42%) (37). Panel (i) shows the total number of guideline violations for each target site (maximum number of violations = 8). There is no correlation between somatic mutation rate and violation of any of these guidelines, as determined using Spearman's rank correlation (see r^2 and P -values on graphs e–i). (j) There is no correlation between TALEN-induced mutation rate and target sequence G+C content ($r^2 = 0.07$, $P = 0.12$ using Pearson's correlation). The linear regression line is shown. Analysis in (e–j) was performed using 34 TALEN pairs.

little or no predictive power for TALEN mutagenicity in zebrafish.

TALEN mutagenicity is negatively correlated with the number of CpG repeats in the target site

It has been shown that binding of TALE domains to their targets is inhibited by 5-methylated cytosine (5mC) and that demethylation of CpG repeats can improve TALE activity in human and rodent cells (42,43). Consistent with these reports, we observed a significant negative correlation between TALEN-induced somatic mutation rates and the number of CpG repeats in target sites (Figure 5b and Supplementary Table S6, $r^2 = 0.29$, $P = 0.001$). TALEN targets containing zero or one CpG repeat exhibited significantly higher mutation rates than those containing two or three CpG repeats (Figure 5b and Supplementary Table S6). In support of this observation, we found that three smaller zebrafish studies that used the same TALEN architecture as our study (9,11,13) showed a similar effect (Supplementary Table S7). In particular, in one study that targeted 10 zebrafish genomic sites, the three targets containing no CpG repeats had the highest mutation rates, whereas the two targets containing five CpG repeats were not mutagenic [Supplementary

Table S7 (13)]. We also analysed data from a large-scale study of TALEN mutagenicity in human cells (41) but found no correlation between mutation rates and the number of CpG repeats in the target (Supplementary Figure S5b and Supplementary Table S8, $r^2 = 0.01$, $P = 0.25$), although targets with no CpG repeats had higher average and median mutation rates than those containing 1–4 CpGs. The basis for the discrepancy between zebrafish and human cells is unknown, but may result from different CpG methylation patterns. In contrast to TALENs, we did not observe a significant correlation between ZFN mutagenicity and the number of CpG repeats in target sites (Figure 5a, Supplementary Figure S5a and Supplementary Table S9, $r^2 = 0.03$, $P = 0.09$). We conclude that, at least for zebrafish studies, TALEN target sites should contain no more than one CpG repeat.

DISCUSSION

ZFNs and TALENs have been used to generate somatic and germ line mutations in zebrafish (3,5–13,15) and relatively small-scale studies have suggested that TALENs may be superior mutagens (9,11–15). However, large-scale comparisons of ZFN- and TALEN-induced mutations have not been reported in any animal. Here, we describe

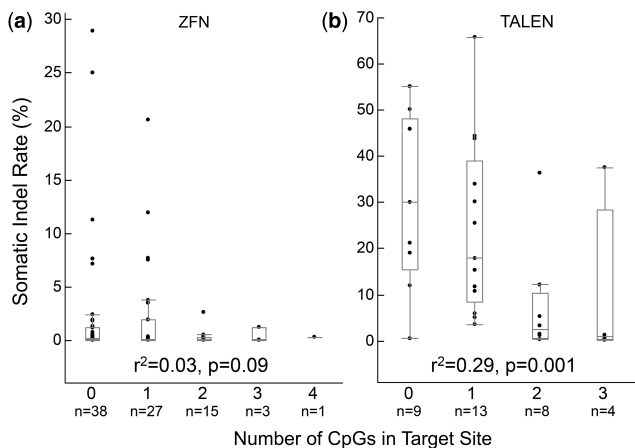


Figure 5. Somatic mutation rate of TALENs but not ZFNs is negatively correlated with the number of CpG repeats in the target site. There is no significant correlation between somatic mutation rate and the number of CpG repeats in ZFN target sequences (a, $r^2 = 0.03$, $P = 0.09$ using Spearman's rank correlation), but there is a significant negative correlation for TALEN target sequences (b, $r^2 = 0.29$, $P = 0.001$ using Spearman's rank correlation). Data points represent somatic indel rates for each of the 84 ZFN (a) and 34 TALEN (b) target sites that we tested. The box within each plot indicates the middle half of the data, and the line in each box indicates the median. The lines extending from the box indicate the farthest data point that is within 1.5 interquartile ranges from the first and third quartiles. Individual data points outside these lines are possible outliers. N indicates number of target sites in a category.

a large-scale, deep sequencing-based comparison of ZFN and TALEN mutagenicity in zebrafish that provides several insights.

First, as has been suggested by other studies (9,11–13), we found that TALENs are significantly more likely to be mutagenic than ZFNs generated using CoDA. Eighty-five percent of the 34 TALEN pairs that we tested induced somatic indels at rates $>1\%$, compared with only 25% of the 84 ZFNs that we tested. Furthermore, all TALENs induced mutations at rates significantly greater than our false-positive rate of 0.01%. We also found that TALENs generated significantly more mutations than ZFNs. The average somatic mutation rates were 20% and 2% for TALENs and ZFNs, respectively, indicating that TALENs are on average 10-fold more mutagenic than ZFNs. Second, by using deep sequencing to screen for somatic mutations, we found that germ line mutants can readily be isolated using ZFNs whose somatic mutation rates are at least as low as 0.27%, which is well below the standard cut-off of 1%, in accordance with observations using a different ZFN methodology (5). If we use a somatic indel rate of 0.27% as a threshold to define nucleases that can generate germ line mutants at reasonable frequencies, 97% of the TALENs and 39% of the ZFNs that we tested are active in zebrafish, with average mutation rates of 21% and 5%, respectively. This analysis also revealed that 89% of ZFNs induced somatic indels at rates above our false-positive rate of 0.01%, suggesting that most ZFNs generated using CoDA are mutagenic, although it may be difficult to isolate germ line mutations using ZFNs that have

somatic indel rates $<0.27\%$. Third, we used our relatively large data set to test whether several guidelines that have been proposed to select optimal TALEN and ZFN target sites are useful predictors of mutagenicity *in vivo* (27,37,40). Our results indicate that none of these guidelines have strong predictive power in zebrafish, and we conclude that they should not be a factor in choosing ZFN and TALEN target sites.

Although published guidelines did not predict ZFN or TALEN mutagenicity *in vivo*, we did find that TALEN mutagenicity is negatively correlated with the number of CpG repeats in the target sequence. TALEN targets containing zero or one CpG repeat exhibited significantly higher mutation rates than those containing two or three CpG repeats. Although not noted in their analysis, three published zebrafish studies that used the same TALEN architecture as our study showed a similar effect (Supplementary Table S7) (9,11,13). These results suggest that target CpG methylation may inhibit TALEN mutagenicity. We note that two TALENs with two or three CpG repeats exhibited high mutation rates in our study (Supplementary Table S6), but it is possible that these sites are not methylated *in vivo*. Consistent with these observations, the presence of 5mC in target DNA can inhibit TALE activity, and demethylation can improve TALE activity, in human and rodent cells (42,43). It has been noted (42) that CpG methylation may be associated with some of the non-mutagenic TALENs described in a large-scale analysis of TALENs in human cells (41), although we found no significant correlation between mutagenicity and the number of target site CpGs in the overall data set (Supplementary Table S8). The cause of the discrepancy between the zebrafish and human cell culture results is unclear, but may result from different CpG methylation patterns in the human cell line that was used compared with developing zebrafish embryos. It has been shown that use of the TALE repeat N* rather than HD at 5mC residues can increase TALEN activity, likely because of reduced steric hindrance of N* compared with HD with the 5mC methyl moiety (42). Taken together, these observations suggest that CpG methylation could be a significant factor in the low mutagenicity of some TALENs, and that targeting CpG residues using N* rather than HD could significantly improve mutagenicity at these targets. Alternatively, we suggest that TALEN targets should not contain more than one CpG repeat, at least for zebrafish studies. In contrast to TALENs, we did not observe a significant correlation between ZFN mutagenicity and the number of target site CpG repeats (Supplementary Table S9).

Studies using plasmid-based and artificial genomic reporters in human cell lines and *Xenopus* oocytes found that ZFNs whose targets contain 5- or 7-bp spacers were less mutagenic than those containing 6-bp spacers (44–46). In contrast, we found that endogenous ZFN targets containing 5- or 6-bp spacers exhibited similar mutation rates, consistent with several studies using animal models (5,8,11). We also found that ZFN targets containing 7-bp spacers were 4- to 5-fold less likely to be mutagenic and had significantly lower mutation rates than targets containing 5- or 6-bp spacers, even though we used ZFN

architectures that were shown to be optimal for 5-, 6- and 7-bp spacers in human cells (44), suggesting that targets containing 7-bp spacers should be avoided.

Our results are consistent with the largest test of TALEN mutagenicity to date, which targeted 96 human genes in cell culture (41). In this study, 88% of TALENs produced mutation rates of 2–56%, with an average of 22%. In comparison, 85% of the TALENs that we tested produced mutation rates of 1–66%, with an average of 20%. Similar to our observations, this study failed to detect correlations between target selection guidelines (37) and mutation rates (41). These results indicate that TALENs mutate endogenous genes at similar rates in human cells and developing zebrafish embryos.

Our results demonstrate that TALENs are highly effective in generating mutations in zebrafish, and that essentially all TALENs are capable of inducing mutations, although mutation rates vary considerably. This variation is likely not completely due to the number of CpG repeats in target sequences, because targets containing the same number of CpG repeats exhibit a wide range of mutation rates (Supplementary Tables S6 and S7), although this might result from different methylation patterns *in vivo*. Further work is therefore needed to understand the basis of this variability, which should be facilitated by high-throughput methods for TALEN construction (41,47,48).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–9 and Supplementary Figures 1–5.

ACKNOWLEDGEMENTS

The authors thank Jason Stein for statistical assistance, Scot Wolfe for sharing sequence alignment algorithms and Catherine Oikonomou for comments on the manuscript.

FUNDING

National Institutes of Health (NIH) [R00 NS060996, R01 NS070911 and R01DA031367 to D.A.P., F31NS077842 to S.C., F31NS077844 to J.L.]; Edward Mallinckrodt Jr., Rita Allen and Brain and Behavior Research Foundations (to D.A.P.); Della Martin Fund (to D.A.P. and G.O.); Millard and Muriel Jacobs Genetics and Genomics Laboratory at the California Institute of Technology. Funding for open access charge: NIH.

Conflict of interest statement. None declared.

REFERENCES

- Urnov,F.D., Rebar,E.J., Holmes,M.C., Zhang,H.S. and Gregory,P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
- Maeder,M.L., Thibodeau-Beganny,S., Osiaik,A., Wright,D.A., Anthony,R.M., Eichtinger,M., Jiang,T., Foley,J.E., Winfrey,R.J.,

- Townsend,J.A. *et al.* (2008) Rapid “open-source” engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell*, **31**, 294–301.
- Meng,X., Noyes,M.B., Zhu,L.J., Lawson,N.D. and Wolfe,S.A. (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 695–701.
- Ramirez,C.L., Foley,J.E., Wright,D.A., Muller-Lerch,F., Rahman,S.H., Cornu,T.I., Winfrey,R.J., Sander,J.D., Fu,F., Townsend,J.A. *et al.* (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods*, **5**, 374–375.
- Zhu,C., Smith,T., McNulty,J., Rayla,A.L., Lakshmanan,A., Siekmann,A.F., Buffardi,M., Meng,X., Shin,J., Padmanabhan,A. *et al.* (2011) Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. *Development*, **138**, 4555–4564.
- Gupta,A., Christensen,R.G., Rayla,A.L., Lakshmanan,A., Stormo,G.D. and Wolfe,S.A. (2012) An optimized two-finger archive for ZFN-mediated gene targeting. *Nat. Methods*, **9**, 588–590.
- Doyon,Y., McCammon,J.M., Miller,J.C., Faraji,F., Ngo,C., Katibah,G.E., Amora,R., Hocking,T.D., Zhang,L., Rebar,E.J. *et al.* (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 702–708.
- Sander,J.D., Dahlborg,E.J., Goodwin,M.J., Cade,L., Zhang,F., Cifuentes,D., Curtin,S.J., Blackburn,J.S., Thibodeau-Beganny,S., Qi,Y. *et al.* (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods*, **8**, 67–69.
- Sander,J.D., Cade,L., Khayter,C., Reyon,D., Peterson,R.T., Joung,J.K. and Yeh,J.R. (2011) Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat. Biotechnol.*, **29**, 697–698.
- Foley,J.E., Yeh,J.R., Maeder,M.L., Reyon,D., Sander,J.D., Peterson,R.T. and Joung,J.K. (2009) Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool Engineering (OPEN). *PLoS One*, **4**, e4348.
- Moore,F.E., Reyon,D., Sander,J.D., Martinez,S.A., Blackburn,J.S., Khayter,C., Ramirez,C.L., Joung,J.K. and Langenau,D.M. (2012) Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). *PLoS One*, **7**, e37877.
- Huang,P., Xiao,A., Zhou,M., Zhu,Z., Lin,S. and Zhang,B. (2011) Heritable gene targeting in zebrafish using customized TALENs. *Nat. Biotechnol.*, **29**, 699–700.
- Cade,L., Reyon,D., Hwang,W.Y., Tsai,S.Q., Patel,S., Khayter,C., Joung,J.K., Sander,J.D., Peterson,R.T. and Yeh,J.R. (2012) Highly efficient generation of heritable zebrafish gene mutations using homo- and heterodimeric TALENs. *Nucleic Acids Res.*, **40**, 8001–8010.
- Dahlem,T.J., Hoshijima,K., Jurynek,M.J., Gunther,D., Starker,C.G., Locke,A.S., Weis,A.M., Voytas,D.F. and Grunwald,D.J. (2012) Simple methods for generating and detecting locus-specific mutations induced with TALENs in the zebrafish genome. *PLoS Genet.*, **8**, e1002861.
- Bedell,V.M., Wang,Y., Campbell,J.M., Poshusta,T.L., Starker,C.G., Krug,Ii,R.G., Tan,W., Penheiter,S.G., Ma,A.C., Leung,A.Y. *et al.* (2012) *In vivo* genome editing using a high-efficiency TALEN system. *Nature*, **491**, 114–118.
- Westerfield,M. (2000) *The Zebrafish Book. A Guide for the Laboratory Use of Zebrafish (Danio rerio)*, 4th edn. University of Oregon Press, Eugene, OR.
- Sander,J.D., Maeder,M.L., Reyon,D., Voytas,D.F., Joung,J.K. and Dobbs,D. (2010) ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res.*, **38**, W462–W468.
- Miller,J.C., Holmes,M.C., Wang,J., Guschin,D.Y., Lee,Y.L., Rupniewski,I., Beausejour,C.M., Waite,A.J., Wang,N.S., Kim,K.A. *et al.* (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.*, **25**, 778–785.
- Foley,J.E., Maeder,M.L., Pearlberg,J., Joung,J.K., Peterson,R.T. and Yeh,J.R. (2009) Targeted mutagenesis in zebrafish using customized zinc-finger nucleases. *Nat. Protoc.*, **4**, 1855–1867.
- Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.

21. Trapnell, C., Pachter, L. and Salzberg, S.L. (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, **25**, 1105–1111.
22. Rumble, S.M., Lacroute, P., Dalca, A.V., Fiume, M., Sidow, A. and Brudno, M. (2009) SHRiMP: accurate mapping of short color-space reads. *PLoS Comput. Biol.*, **5**, e1000386.
23. David, M., Dzamba, M., Lister, D., Ilie, L. and Brudno, M. (2011) SHRiMP2: sensitive yet practical SHort Read Mapping. *Bioinformatics*, **27**, 1011–1012.
24. Kent, W.J. (2002) BLAT—the BLAST-like alignment tool. *Genome Res.*, **12**, 656–664.
25. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
26. Doyon, Y., Vo, T.D., Mendel, M.C., Greenberg, S.G., Wang, J., Xia, D.F., Miller, J.C., Urnov, F.D., Gregory, P.D. and Holmes, M.C. (2011) Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat. Methods*, **8**, 74–79.
27. Sander, J.D., Reyon, D., Maeder, M.L., Foley, J.E., Thibodeau-Beganny, S., Li, X., Regan, M.R., Dahlborg, E.J., Goodwin, M.J., Fu, F. *et al.* (2010) Predicting success of oligomerized pool engineering (OPEN) for zinc finger target site sequences. *BMC Bioinformatics*, **11**, 543.
28. Zou, J., Maeder, M.L., Mali, P., Pruetz-Miller, S.M., Thibodeau-Beganny, S., Chou, B.K., Chen, G., Ye, Z., Park, I.H., Daley, G.Q. *et al.* (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell*, **5**, 97–110.
29. Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K. and Voytas, D.F. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature*, **459**, 442–445.
30. Zhang, F., Maeder, M.L., Unger-Wallace, E., Hoshaw, J.P., Reyon, D., Christian, M., Li, X., Pierick, C.J., Dobbs, D., Peterson, T. *et al.* (2010) High frequency targeted mutagenesis in *Arabidopsis thaliana* using zinc finger nucleases. *Proc. Natl Acad. Sci. USA*, **107**, 12028–12033.
31. Bae, K.H., Kwon, Y.D., Shin, H.C., Hwang, M.S., Ryu, E.H., Park, K.S., Yang, H.Y., Lee, D.K., Lee, Y., Park, J. *et al.* (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat. Biotechnol.*, **21**, 275–280.
32. Carroll, D., Morton, J.J., Beumer, K.J. and Segal, D.J. (2006) Design, construction and in vitro testing of zinc finger nucleases. *Nat. Protoc.*, **1**, 1329–1341.
33. Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. and Kim, J.S. (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.*, **19**, 1279–1288.
34. Liu, J. and Stormo, G.D. (2008) Context-dependent DNA recognition code for C2H2 zinc-finger transcription factors. *Bioinformatics*, **24**, 1850–1857.
35. Sander, J.D., Zaback, P., Joung, J.K., Voytas, D.F. and Dobbs, D. (2009) An affinity-based scoring scheme for predicting DNA-binding activities of modularly assembled zinc-finger proteins. *Nucleic Acids Res.*, **37**, 506–515.
36. Segal, D.J., Dreier, B., Beerli, R.R. and Barbas, C.F. 3rd (1999) Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl Acad. Sci. USA*, **96**, 2758–2763.
37. Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.
38. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
39. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
40. Streubel, J., Blucher, C., Landgraf, A. and Boch, J. (2012) TAL effector RVD specificities and efficiencies. *Nat. Biotechnol.*, **30**, 593–595.
41. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.*, **30**, 460–465.
42. Valton, J., Dupuy, A., Daboussi, F., Thomas, S., Marechal, A., Macmaster, R., Melliand, K., Juillerat, A. and Duchateau, P. (2012) Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. *J. Biol. Chem.*, **287**, 38427–38432.
43. Bultmann, S., Morbitzer, R., Schmidt, C.S., Thanisch, K., Spada, F., Elsaesser, J., Lahaye, T. and Leonhardt, H. (2012) Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. *Nucleic Acids Res.*, **40**, 5368–5377.
44. Handel, E.M., Alwin, S. and Cathomen, T. (2009) Expanding or restricting the target site repertoire of zinc-finger nucleases: the inter-domain linker as a major determinant of target site selectivity. *Mol. Ther.*, **17**, 104–111.
45. Shimizu, Y., Bhakta, M.S. and Segal, D.J. (2009) Restricted spacer tolerance of a zinc finger nuclease with a six amino acid linker. *Bioorg. Med. Chem. Lett.*, **19**, 3970–3972.
46. Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.G. and Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell Biol.*, **21**, 289–297.
47. Wang, Z., Li, J., Huang, H., Wang, G., Jiang, M., Yin, S., Sun, C., Zhang, H., Zhuang, F. and Xi, J.J. (2012) An integrated chip for the high-throughput synthesis of transcription activator-like effectors. *Angew. Chem. Int. Ed. Engl.*, **51**, 8505–8508.
48. Briggs, A.W., Rios, X., Chari, R., Yang, L., Zhang, F., Mali, P. and Church, G.M. (2012) Iterative capped assembly: rapid and scalable synthesis of repeat-module DNA such as TAL effectors from individual monomers. *Nucleic Acids Res.*, **40**, e117.